Acing the Test: Developing a Test Committee
Janice Tazbir RN, MS, CS, CCRN, CNE, RYT
Janet Landrum RN, MS, CNS
Cynthia Cooke RN, MS, CCRN

Learning Objectives
1. Summarize testing committee best practices
2. Recognize one CON testing committee’s experience
3. Recognize the testing committee role as faculty resource for test development and analysis
4. Use information learned to create or refine a testing committee

Description
• Testing is an incredibly important part of nursing education. As NCLEX passing rates decline, faculty are trying to find ways to better prepare students for the NCLEX exam. A testing committee is an active, encompassing way to address testing in any curriculum. Testing committees require conversation and consensus as well as a team approach. This presentation will share the experience of one undergraduate college of nursing in their creation of a testing committee.
Research and Best Practices

- Schroeder
- Application level, rationales, multilogical thinking, high level of discrimination, test analysis, testing manual, testing as NLNAC systematic plan of ongoing assessment of student learning and curriculum evaluation (internal and external)
- NLN White paper
- Ethical obligation to fair and valid testing, selection of appropriate tests, communication plan, administration of tests and scoring, fair testing environment, testing policy

- Evolve Elsevier White Paper
- Communicate expectations, + testing culture, reinforce testing and academic dishonesty policies, testing consequences, remediation, data analysis
- Barton
- Benchmarks, test prep, remediation, testing policies translate to higher HESI exit score
Testing policies associated with higher HESI Exit Score

<table>
<thead>
<tr>
<th>Policy</th>
<th>With Policy</th>
<th>Without Policy</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandatory benchmark</td>
<td>967.2 (112.90)</td>
<td>895.16 (124.27)</td>
<td>t = -4.435, p = .001</td>
</tr>
<tr>
<td>Required participation to pass</td>
<td>905.47 (112.83)</td>
<td>905.43 (124.14)</td>
<td>t = -0.071, p = .944</td>
</tr>
<tr>
<td>Required to re-take</td>
<td>898.76 (114.83)</td>
<td>898.98 (128.47)</td>
<td>t = -0.074, p = .944</td>
</tr>
<tr>
<td>Required remediation</td>
<td>885.15 (111.55)</td>
<td>840.95 (127.65)</td>
<td>t = -4.011, p = .001</td>
</tr>
</tbody>
</table>


Why make a Testing Committee

• Develop consistency in student and curriculum evaluation
  • guidelines for internal evaluation (statistical parameters such as level of difficulty, level on Bloom's and other criteria for faculty made exams)
  • external curriculum evaluation (compares students to overall population on exams and NCLEX pass rate).
• Develop consistency in testing throughout program
  • rules and expectations on test taking and how tests will be used in the curriculum are clear for students and faculty (i.e. test types, timing, development criteria for faculty made exams)

Why make a Testing Committee? Don’t we have enough to do?

• Testing Writing
• Test Analysis
• Consistency of faculty in testing
• Fairness and obligation we have to students
• Keeping current with test plan and item writing
How a Testing Committee Helps Students

- Prepare students with practice questions written at the analysis level and above.
- Establish positive testing culture.
- Establish clear expectations for student achievement
- Build Student Confidence - discuss testing techniques and strategies for student success

Role of a Testing Committee

- Develops, implements, and evaluates the testing policy (includes faculty made tests and standardized testing program).
- This committee must be supported by administration.
- Reviews current literature on testing, establishing benchmarks, etc.
- Recommends revisions for the test policy

What are the “big chucks”?  

- Learn more about the NCLEX exam
- Translate clinical thinking into NCLEX style questions
- Test writing and interpretation
- Create common language and standards as it relates to testing (testing committee)
- Student success!!
Committee Objectives:

- Review and revise CON testing to reflect best NCLEX style testing practice in the undergraduate curriculum.
- Create a schedule for exam submission, committee review with recommendations, and return to faculty for revisions and resubmission.
- Support ongoing faculty development in test construction and analysis.

Charge

- **Testing Committee Charge:**
  - The testing committee consists of undergraduate course representation from the CON faculty. The committee is charged to guide and support faculty in creating, implementing and analyzing tests and remediation based on the current NCLEX test plan and testing best practices.
- **Functions of the Testing Committee:**
  - Review and revise CON testing to reflect best NCLEX style testing practice in the undergraduate curriculum.
  - Create a schedule for exam submission, committee review with recommendations, and return to faculty for revisions and resubmission.
  - Support ongoing faculty development in test construction and analysis.
Request Letter

Dear Colleague,

The testing committee was formed to provide support to the CON for test development and analysis for the nursing courses. The testing committee respectfully requests to review your course exam(s). The materials needed by the committee include:

- Statistical analysis of your exam, Question level (according to Bloom’s taxonomy)
- KR20, Total percent correct, Point biserial, Distractors, Question level

If you want further help in actual test question creation and refinement, the committee will be available on an as needed basis.

To assist you in understanding the feedback provided, see the attached documents:

- Sample statistical analysis of an example
- Sample test blueprint
- Example of feedback provided (on the statistical analysis sheet)
- Explanation of statistics sheet.
- Taxonomy sheet (see site: [http://www.celt.iastate.edu/teaching-resources/effective-practice/revised-blooms-taxonomy/](http://www.celt.iastate.edu/teaching-resources/effective-practice/revised-blooms-taxonomy/))

Thank you for helping the CON in providing excellence in nursing education.

Testing Policy

Principles

Faculty provide the students with the content, tools, resources, remediation opportunities and critical thinking skills to become effective test takers and to pass the NCLEX examination. Faculty have a responsibility in providing well written exams that reflect NCLEX testing standards which are reviewed using statistical standards. Testing is used as a form of internal and external curriculum evaluation as a measure of student learning outcomes.

Beyond course grades, successful testing is necessary to pass the NCLEX-RN exam and ultimately become an RN. Students have a responsibility to be accountable for all preparation, remediation, actions, behaviors and consequences relating to testing.

Students have a responsibility to uphold the ANA nursing code of ethics, policies and statements related to testing in the undergraduate handbook, and university policies related to academic honesty and the honor code. Testing related items may be addressed in specific course syllabi in addition this policy.

Guidelines

The College of Nursing testing policy is introduced to all new students at orientation and is included in the undergraduate handbook. In addition, each course syllabus outlines the protocol for evaluation of student learning and minimum passing standards. Students should be aware of the importance of scoring well on both faculty developed and standardized examinations, and the remediation resources and retesting opportunities available.

Testing Behaviors

The majority of nursing courses have periodic examinations throughout the semester and specifics related to the exams will be addressed in each course. Whether the exams are paper and pencil or computer based:

- Academic integrity and the honor code should be explained, reinforced and upheld.
- Consequences associated with academic dishonesty are described in the testing policy, the undergraduate handbook, course syllabi. Those students found to be in violation are referred to the office of the Dean of Students.
- Time allowances will be set in advance and adherence to time will be enforced.
- It is the responsibility of a student with an ADA accommodation who requires additional time for testing to notify the course instructor at the beginning of the semester or as soon as the disability is documented through the Office of Disability Resources.
- Personal items, including phones, backpacks, bag, purses and books will be secured and not accessed by any student during the exam.
- Students may be excused to leave the room in case of illness or need to use the restrooms. The exam will be secured by the proctor or faculty member during the student’s absence.
- Students may be asked to remove hats, caps, hoodies, and specific jewelry prior to exams. Failure to comply may result in the student forfeiting the exam and receiving a “0”.
- If an exam requires audio, students will be responsible for bringing their own earbuds, none will be provided by a proctor/faculty member.
All students are REQUIRED to have their PUC ID in their possession as an ID verification for all exams. Failure to comply will result in the student forfeiting the exam and receiving a "0".

If a student is caught cheating or suspected of cheating in the testing center, the proctor will notify the faculty member and the student will be contacted by the faculty member regarding the incident and dealt with accordingly.

When testing in the testing center, it is the student’s responsibility to follow all testing center regulations including no food or drinks, sign in procedures and ID verification for all testing. When allowed, students may be issued a blank piece of paper for “scrap paper”. All papers will be collected by the faculty member or proctor at the end of the exam.

Testing rooms are expected to be quiet and the use of noise reducing earplugs is considered acceptable.

Calculators use will be described in each course. When computer testing, the calculator will be accessed from the computer. In classrooms, calculators will be available.

Post Exam
The following should be employed after an exam:
Students will have grades posted after exam analysis is complete.
Students will have an opportunity to review exams and rationales.
Students will be offered remediation activities to assist in content mastery and test taking skills.

Testing 101

• 2013 NCLEX test plan
• Test plan vs. “home grown”
• Question style
• Stem examples
• Blueprints
• Test and item analysis
• CON Testing consensus/best practices

Testing Blueprint

<table>
<thead>
<tr>
<th>Question</th>
<th>Nursing Process</th>
<th>Client need</th>
<th>Cognitive Level</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Analysis

- Develop a plan for evaluating test data. How often should data be analyzed? What trends should be analyzed routinely? Determine the key parameters to assess.
- Review trends over time and not just one data point.
- Be sure to not only analyze data but develop an action plan based on findings as well. What will you do with the data? Need an action plan - for example identify 3 areas to focus on with next student cohort.

Discussion

Types of things to discuss in a testing committee:
- Test length
- Question level
- Environment
- Proctoring
- Question types
- Remediation

Testing Committee

- Ad hoc vs. standing
- Composition
- Responsibilities
A measurement tool, designed to measure the learners’ knowledge, skills and abilities.

Defining a Test Item

**Constructed-Response Item**
- Instruction or question is written
- Student gives a response
- Teacher develops a scoring system
- Teacher uses scoring system

**Selected-Response Item**
- Instruction or question is written
- Teacher constructs choices
- Student selects among choices
- Objective scoring is used

Basic *Selected-Response* Test Construction

1. Determine purpose
2. Define underlying theory
3. Develop a test plan
4. Develop items
5. Build the test form
6. Standardize testing conditions
7. Establish grading criteria
8. Report and analyze
Objectives Drive Assessment Strategies

Effective instruction is based on the identification of content that is worthy of instruction and the communication of this instructional intent to the students.

Haladyna

---

Item Types

- Multiple Choice
- Multiple Response
- Fill-in-the-blank
- Ordered Response
- Hot spot
- Charts/Exhibits

---

Item Template

Choose the type of student outcome

Knowledge | Mental Skill | Mental Ability

What content are you teaching/measuring?

Fact | Concept | Principle | Procedure

What type of mental behavior are you developing?

Recall | Understand | Apply/Analyze | Create

What format will you use?

MC | MR | OR | Drag/Drop
Process of examining class-wide performance on individual test items

Types of Item Analysis

- **Item Difficulty Index, P**
- **Item Discrimination, R(IT)**
- **Test Reliability**

**Difficulty Index, P**

- Percentage of students that correctly answered the item
- p-value
- Range from 0.0 to 1.00 (representing 0% to 100%)
- The higher the value, the easier the item.
- P-values > 0.90 are very easy
- P-values < 0.30 are very difficult - nullify.
- Optimal difficulty level is 0.50.
**Difficulty Index**

<table>
<thead>
<tr>
<th>Item</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0</td>
<td>3</td>
<td>24*</td>
<td>3</td>
</tr>
<tr>
<td>#2</td>
<td>12*</td>
<td>13</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

* Denotes correct answer

\[ P = \frac{\text{# of persons with item correct}}{\text{# of persons who took the test}} \]

---

**Discrimination Index, R(IT)**

Refers to how well an assessment differentiates between high and low scorers.

- Also referred to as the Point-Biserial correlation (PBS)
- Range from 0.0 to 1.00.
- The higher the value, the more discriminating the item.
- Items with low discrimination values (near or less than zero) should be revised or removed from the test.

---

**Classroom Test Discrimination Guideline**

- Greater than 0.30: Good items
- 0.20 to 0.29: Reasonably good items
- 0.09 or 0.19: Fair (acceptable) items
- Less than 0.09: Review and possibly nullify
Item Discrimination

<table>
<thead>
<tr>
<th>Item #1</th>
<th>Correct</th>
<th>% Choosing</th>
<th>Difficulty</th>
<th>PBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item  #1</td>
<td>A Key</td>
<td>0.72</td>
<td>0.28</td>
<td>0.40</td>
</tr>
<tr>
<td>Item  #1</td>
<td>B</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item  #1</td>
<td>C</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item  #1</td>
<td>D</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item  #1</td>
<td>E</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item #2</th>
<th>Correct</th>
<th>% Choosing</th>
<th>Difficulty</th>
<th>PBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item  #2</td>
<td>A Key</td>
<td>0.72</td>
<td>0.28</td>
<td>0.04</td>
</tr>
<tr>
<td>Item  #2</td>
<td>B</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item  #2</td>
<td>C</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item  #2</td>
<td>D</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item  #2</td>
<td>E</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Item Discrimination

<table>
<thead>
<tr>
<th>Item #3</th>
<th>Correct</th>
<th>% Choosing</th>
<th>Difficulty</th>
<th>PBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item  #3</td>
<td>A</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item  #3</td>
<td>B</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item  #3</td>
<td>C</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item  #3</td>
<td>D Key</td>
<td>0.98</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Item  #3</td>
<td>E</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item #4</th>
<th>Correct</th>
<th>% Choosing</th>
<th>Difficulty</th>
<th>PBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item  #4</td>
<td>A</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item  #4</td>
<td>B</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item  #4</td>
<td>C</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item  #4</td>
<td>D Key</td>
<td>0.09</td>
<td>0.30</td>
<td>-0.19</td>
</tr>
<tr>
<td>Item  #4</td>
<td>E</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bloom’s Taxonomy and Item Discrimination

By categorizing items according to Bloom’s Taxonomy, one may determine if the difficulty index and the discrimination index of these groups of questions are appropriate.
Test Reliability

Two measures for test reliability
• Kuder-Richardson 20 (KR-20)
• Coefficient Alpha

Test Reliability

• Range from 0.0 to 1.0
• The higher the value, the more reliable the overall test score.
• Measure of internal consistency reliability
• High reliability indicates that the items are all measuring the same thing.

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90 and above</td>
<td>Excellent reliability; at the level of the best standardized tests</td>
</tr>
<tr>
<td>0.80 – 0.90</td>
<td>Very good for a classroom test</td>
</tr>
<tr>
<td>0.70 – 0.80</td>
<td>Good for a classroom test; in the range of most. There are probably a few items which could be improved</td>
</tr>
<tr>
<td>0.60 – 0.70</td>
<td>Somewhat low. This test needs to be supplemented by other measures to determine grades. There are probably some items which could be improved.</td>
</tr>
<tr>
<td>0.50 – 0.60</td>
<td>Suggests need for revision of test, unless it is quite short (ten or fewer items). The test definitely needs to be supplemented by other measures for grading.</td>
</tr>
<tr>
<td>0.50 or below</td>
<td>Questionable reliability. This test should not contribute heavily to the course grade, and it needs revision.</td>
</tr>
</tbody>
</table>
Improving Test Reliability

- Increase the number of questions in the test.
- Use items that have high discrimination values.

Distracter Evaluation

- Relationship between the distracters students choose and total test score
- Quality of distracters influences student performance on a test
- Distracters should appeal to low scorers who have not mastered the material
- Distracters can be revised, replaced or removed.

A frequency table provides a means to study responses to distracters.

So now what….how to begin

- Form a committee (who should be on it and why)
- Is there a resident expert?
- Review the research
- Explore present testing (where are we and where do we have to go)
- What are the goals?
- Where are the issues?
- Educate all faculty
Lessons Learned

• Test expert
• Involve faculty
• Make it easy on the ego
• Let the statistics do the talking
• Explore testing software
• Testing center
• Communicate
• Educate
• Revise/Revisit

References


Elsvevier. Testing Best Practices (White Paper)


Thank you for your time and attention!